PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The effects of health status on work exit and absenteeism among the older working population in China
AUTHORS	Li, Xin; Zhang, Wei; Sun, Huiying; Anis, Aslam

VERSION 1 - REVIEW

REVIEWER	Stephen Stansfeld Queen Mary University of London, United Kingdom
REVIEW RETURNED	22-May-2018

GENERAL COMMENTS	This is an interesting paper studying work exit and absenteeism in rural and urban populations in relation to self-rated health status in China. Title and Abstract: 'Work productivity loss' is somewhat ambiguous. You could consider calling it work exit and absenteeism instead. Page 5, line 40: can you define a 'cadre' please? Page 7, line 12: What were your hypotheses? Page 7, line 28: Some detail of the sampling frame, response rate, geographical area from which the study population was drawn would be helpful here without having to refer to Zhao et al. Page 8, line 54: Can you give more details in the main text of how these other health conditions were measured. Did you use standard scales? The discussion could include more comparison with the results of international studies.

REVIEWER	Jan Fekke Ybema Utrecht University, The Netherlands
REVIEW RETURNED	20-Nov-2018

GENERAL COMMENTS	This is an interesting manuscript about the relationship between self-report health and exit from work and sickness absence among older workers in China. The manuscript reports on a longitudinal study with a baseline measurement in 2011, and a follow-up measurement 2 years later in 2013. It focuses on changes in health status (poor vs. good/fair) and how this is related to exit from work between both measurements and sickness absence as reported at follow-up. It distinguished four groups of workers: male

	and female farmers and non-farmers. This manuscript seems to be a revised manuscript, however, it is new to me, so I have little choice but to look at it as a new manuscript. Although the manuscript has a number of interesting features, I also see several serious flaws and limitations that I think should addressed before the manuscript would be ready for publication. The most interesting feature is the attempt to find predictors of exit from work among an understudied group of workers in China. This is highly relevant given the aging population. I would like to focus on the following problematic aspects of the manuscript that I think need to be changed or addressed: (1) the analytical strategy; (2) the reasons for work exit; (3) the validation of self-reported health; (4) the role of demographic and other control-variables; (5) the term “productivity loss”; (6) the role of sickness absence; (7) details of data collection. (1) Analytical strategy The most important point is the analytical strategy. Now, in the central analysis on which conclusions are based, the change in self-reported health (SRH) in 2011 and in 2013 is used to predict exit from work between 2011 and 2013. This means that SRH in 2013 is measured AFTER exit from work. In longitudinal research, independent variables (SRH) should be measured BEFORE the dependent variable (exit). This means that only the results for model I are based on prospective analyses. I think the results for model II (“concurrent health” but actually health following exit), model III, and model IV should be deleted. This simplifies the manuscript strongly, and the resulting analysis is truly prospective. It might be interesting to examine SRH at 2013 as a DEPENDENT variable that is regressed on SRH 2011 and on exit between 2011-2013. This would indicate to what extent exit from work lowers or increases SRH. However, this would be the answer to a different research question than is currently examined. Secondly, I think the separate analyses for males and females, and for the four groups (female farmers, female non-farmers, male farmers, male non-farmers) should ONLY be presented when gender, occupation and their interaction significantly INTERACT with SRH 2011 in the regression of exit from work. If this is not the case, gender and/or occupation (and their interaction) could just be entered as main effects in the regression of exit from work. This would probably be better, given the relatively low numbers of exit from work in some of the groups. My guess would be that there will only be main effects for gender (females have a higher risk of exit) and occupation (non-farmers have a higher risk of exit), in addition to a main effect of SRH 2011 (more exit as health is poorer). Thirdly, I do not think it is necessary to dichotomize or trichotomize SRH. The full 5-point scale could be used as a “continuous” (or preferably ordinal) scale. Fourthly, I do like trajectory analyses in which changes in SRH are related to the timing of exit from work. However, to do such trajectory analyses in a useful way, you would need more than two measurements. And preferably a much more fine-grained measurement, in which for example each month SRH is measured

	and changes in SRH before and after the moment of exit from work are analysed using a time series approach. Finally, it is not fully clear to me what the benefits are of the relatively complex PROBIT analysis over a normal logistic regression, and how the results would be different. I understand that the focal regression (of work exit) is simultaneously done with the validation of SRH. However, I do not understand how this would change the results of either of these analyses. I also think the explanation in the Appendix is now overly technical, and the more interesting results from these analyses are missing (see below, (3) and (4)). (2) Reasons for work exit As is mentioned in the Discussion, a minority of the participants who exit from work were not working due to disability or retirement. What other reasons did participant have for not working? I suppose that many of them were unemployed or that there was no work available. We know that unemployment leads to health problems (e.g., the meta-analysis by McKee-Ryan et al., JAP, 2005), which makes it even more important to distinguish health as a predictor and health as a result of exit (see (1)). Is it possible to distinguish voluntary and non-voluntary exit from work? It may be interesting to distinguish several exit routes, as was done in Leijten, et al (2015, your reference 18), using competing risk proportional hazard models. (3) Validation of SRH In the Appendix an econometric model is specified. If I understand correctly, equations (3) and (4) regard some kind of validation of SRH based on more detailed health measures. However, no results of this validation is given. What percentage of variance of SRH is explained by these detailed measures? Would it be possible to create a more valid and detailed single health measure based on these scales? The items strongly resemble the physical items of SF-12 or SF-36, which are used to create validated scores for mental health and a score for physical health on a scale from 0 to 100. Such a score would seem more reliable and valid than the trichotomized single item on SRH that is used in the analyses. (4) Demographic and other control variables Apparently, in the regression of exit from work several control variables are included (given equation 1 in the Appendix, page 33). I would like to see the results of these control variables. How were age, education, marriage status, and expenditures related to exit from work? Moreover, did any of these variables moderate the relationship between SRH 2011 and work exit? (5) The term “productivity loss” I think the term productivity loss is confusing and should be removed from the title, and perhaps throughout the manuscript. I would advise the authors to use the term “exit from work” instead (and perhaps “sickness absence”, but see next point). Productivity loss is generally reserved to loss of productivity while working (i.e., due to presenteeism) as a complement of sickness absence. (6) The role of sickness absence I think it is not that interesting to examine the relationships between SRH and sickness absence. Sickness absence is

	evidently related to lower health (otherwise it would not be sickness absence). This relationship would only be interesting if it is moderated in a meaningful way by other variables. This may be the case for occupation (farmer vs. non-farmer), but this should be tested first. More interesting would be moderation by working conditions (e.g. job demands or job resources that would amplify or buffer the relationship between health and sickness absence). However, as it is presented now, I do not think the inclusion of sickness absence is worthwhile. Moreover, I think the tentative conclusion that for farmers sickness absence is a way to deal with the lack of opportunities for retirement is highly speculative and not based on the results of the analyses. (7) Data collection Finally, I would like some more information about how the data were collected. Were all participants interviewed face-to-face? If not, how did illiterate participants respond? Despite these problems that I see in the current manuscript, I would strongly suggest the author to substantially rework the analyses and revise the manuscript. I do believe this could result in a valuable contribution to the literature on exit from work.

REVIEWER	Emmanuel Aboagye Karolinska Institutet, Sweden
REVIEW RETURNED	04-Dec-2018

GENERAL COMMENTS	Reviewer's Evaluation: The effects of health status on work productivity loss among the older working population in China Comments to authors The study examines the effects of health status on work productivity loss. It presents estimates of productivity loss (work exit and absenteeism) due to health problems measured as self-rated health (SRH) among older working people in China. Authors conclude that poor lagged health status or poor concurrent health status is expected lead to work exit and more absent workdays. Also persistent poor health status over time is detrimental to the labour participation of the older working population except for female non-farmers. The effects of health status differ by both gender and type of work. Tailored policies to these different working groups are needed. Title Change of title suggestion: The effects of health status on work exit and absenteeism among the older working population in China Work exit or number of absent workdays due to health problems is not necessarily synonymous with productivity loss. Since the study outcomes are work exit and number of absent workdays, it may be advisable to have it in the title. NB: that the assumption of within-subject changes can also be explained that a high performing worker who becomes less productive because of health problems might still be one of the best among peers in terms of global performance. What about replacement and what about compensatory strategies by colleagues or when the sick absent worker comes back to work? Abstract

	1. Work productivity loss was measured in two ways: work exit and the number of absent workdays. This is just a passing comment. I will return to this in the methods and discussion section. 2. Health problems/ status: It will be good to mention what health problems or health status is and how it was measured in the methods of the abstract. Introduction 1. There is a clearly stated aim for the study. It may help that the aim of the study stands alone. Do authors refer to labour force participation and work exit as the same thing? I guess so. You may want to be consistent with the use of the terminologies throughout the manuscript. Please check also the use of labour market activities and continuous working status. 2. Lines 12-20 on page 6 can be moved to part of the introduction which tells readers about other specific aims. Specifically, a further aim of the study was to examine the impact of previous health status and concurrent health status as well as the change of health status over time on work exit and the number of absent workdays due to the health problems among the older people who were previously working. 3. I think it will help to provide a reference for this sentence on line 33-38 on page 5: Most studies to date have focused on either comparing the incremental effects of different chronic diseases on absent workdays or estimating the incremental productivity loss due to different chronic diseases (REF). 4. Data source belong in the methods: It was mentioned that the study uses longitudinal, individual level data from the China Health and Retirement Longitudinal Study (CHARLS). I think all this belong in the method section. Methods 1. Based on the previous comment, I know you have defined productivity loss as work exit and the number of absent workdays but thinking of farmers makes me wonder if it is indeed a loss if a farmer for instance is absent some days. Can the authors clarify why were particularly interested in this classification of agricultural work and also comparing them with non-agricultural work? For absent workdays, how were those working in agriculture respond since they may not be expected to be on the farm every day? 2. Lines 24-31 on page 7 belong to the section where you describe the procedure for analysis. To investigate the potential worker heterogeneity in the effects of health status on work productivity loss, we divided our sample into four separate groups according to gender and working types in 2011: female farmers (i.e., any agricultural work), female non-farmers (nonagricultural work only), male farmers, and male non-farmers. 3. On line 43-45 on page 7 it is said that Labour force participation status was determined by a series of questions in CHARLS. Can authors give an example or two of the nature of the questions used? I am also curious about what questions they have used to measure the different types of labor participation including household agricultural work, temporary work, own business and unpaid work etc. 4. Same concern here as in comment 3. Please check line 54 on page 8. These measures included disability condition, number of chronic diseases, and functional limitations. Please provide examples of the questions asked in the appendix or rather still use in-text references to the detailed health questions the first time

	they are mentioned so that readers can check what kind of question were asked. 5. Can authors provide any ethics statement? State, if it is perhaps not applicable in this case. Results and discussion 1. I think the results and discussion sections address the research questions and cover many grounds including the sensitivity analysis performed, and in conclusion where policy measures may be necessary. 2. It seems that female non farmers clearly stand out in terms of change of health status over time on the outcomes - work exit and the average number of absent workdays due to health problems. What could be the explanation for this? I think the authors can elaborate on this in the discussion beyond the fact that there were only a few female non farmers. Appendix 2. Econometric model specifications We used the Model III, including both lagged and concurrent health status independently he model, as an example to show our model specifications. Please check line 5 on page 33. There is some text missing where I have marked.

REVIEWER	Zhiyuan Hou Fudan University, China
REVIEW RETURNED	18-Dec-2018

GENERAL COMMENTS	This paper addresses an important question. It contributes to literature by estimating the effects of health status on work productivity among the older working population in China, and the results are presented clearly. However, there are several technical details I prefer be revised. First, the paper uses work exit or the number of absent workdays to measure productivity loss. However, in literature, productivity is often measured as TFP or output related variables such as output per worker. Second, analyses in this paper are based upon regression results of 4 separate groups. However, estimates by groups do not provide us evidence about the significance of the difference between two groups. Another way to show heterogeneous effects of health status on productivity for different groups is to use interaction terms in regressions. For example, you can interact gender with health to show if health has significantly different effects on work exit. Please explain why you choose regression for separate groups instead of interaction effect. Third, though instruments are used to estimate the model, regression results suffer from endogeneity issues. The assumption that error terms in the first stage and the second stage are not correlated is very strong. For example, city/regional level characteristics such as economic growth or labor market demand shocks may influence residents' health status and probability of work exit simultaneously, resulting in biased estimates. More city/regional level explanatory variables related to economic/labor market conditions should be included in regressions. Besides, people may exit the labor market due to different reasons. On the one hand, people may exit the job market

	voluntarily due to bad health conditions. On the other hand, people may become unemployed (or exit the labor market involuntarily) due to negative demand shocks in the labor market. This paper uses work exit as the dependent variable for both situations, making the interpretations of the results quite confusing. Fourth, I suggest the authors show how probability of not working, or the expected number of absent workdays are computed in Table 3 to Table 5. It is true that estimated parameters measure the effect of independent variables on latent variables instead of actual dependent variables. Partial effects can always be calculated using maximum likelihood estimations. Please provide marginal effects of health status on work exit and absent workdays. Finally, the paper is policy oriented and the question investigated is very important. I hope more policy discussions could be provided.

REVIEWER	Cathy Gong Australian National University
REVIEW RETURNED	19-Dec-2018

GENERAL COMMENTS	Review comments on BMJ Open-2018-024115 Title: The effects of health status on work productivity loss among the older working population in China Suggestion: Major revision Summary: This paper used the first two-wave longitudinal data from the China Health and Retirement Longitudinal Study (CHARLS) to analyze the effects of health status and its changes on work productivity loss among the older working population in China. It measured the impact of self-reported health on work exit and the number of absent workdays, for four groups according to gender and work types (farmers who conducted any agricultural work at the first wave versus non-farmers who conducted non-agricultural work only). Strength This paper has filled in the research gap to address a very important policy issue on how to encourage older people to work as long as they can in order to face the emerging workforce shortages due to the fast population ageing by improving older people's health. The use of measures of work exit and absent days from work, previous health status, current health status and health changes across two waves are very `interesting and meaningful. It is expected that persistent poor health over time has larger impact on work exit and absence days of work. Limitations:

	(1) The statistical methods used are not enough to support the research question, which need to be improved. (2) The interpretation of stories by gender, famers/non-farmers is not deep enough. Major comments: (1) How is the quality of instrumental variables (IV) (IV=disability condition, number of chronic diseases, and functional limitations) used to address possible endogeneity between self-reported health and work exit or absence days from work? Can you say the three IV variables had direct impact on self-reported health but not on work exit and absent workdays? (2) In you case, as you have two waves of longitudinal data, you might be able to just use difference-in-difference methods or conditional difference-in-difference methods to control for the endogeneity between work and health, instead of IV method. Other comments in details: (1) As previous and current health status are highly correlated, you should use either initial health plus health changes, or current health plus health changes in the models of Tables 3 and 4 , but not using previous and current health status in Table 3 or health changes only in Table 4. (2) You need provide data source and notes for all the tables; (3) It is important to check whether using weights or not generated robustness results, but there is no need to report all the tables in the appendix. Instead, I would like you to report all the estimated coefficients from your models on self-reported health, work exit and absent workdays in the appendix. (4) There are a lot of important differences between farmers and non-farmers, males and females in Table 1 and Table 2, which are very interesting, but you have not discussed them. For instance, non-farmers have much better health than farmers, while famers work longer to late life than non-farmers; changes from poor to poor health has large impact for female non-farmers, male farmers, male non-farmers, while changes from good/fair to poor health has large impact on female farmers, is this because female farmers take more role in unpaid domestic work instead of paid agriculture work? Why worsening health has larger impact on work exit for male non-farmers than female non-farmers, is this because relatively later retirement age for males? Should females have same retirement age with males from health view? (5) Will your findings indicate that retirement age could be be postponed to later age in urban area from health view? Will work flexibility (like in rural area) help older people to work longer by taking more unpaid leave while staying in workforce longer? (6) You have checked the robustness of your results by removing the age restriction for farmers, while it will be more interesting to compare the impact of health changes for farmers before and after legal retirement age. This could be used to inform non-farmers behaviours if the retirement age is cancelled. (7) Do you know how many people are still working after retirement age for farmers and non-farmers?

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1
Reviewer Name: Stephen Stansfeld
Institution and Country: Queen Mary University of London, - United Kingdom
Please state any competing interests or state 'None declared': none declared

Please leave your comments for the authors below
This is an interesting paper studying work exit and absenteeism in rural and urban populations in relation to self-rated health status in China.
Title and Abstract: 'Work productivity loss' is somewhat ambiguous. You could consider calling it work exit and absenteeism instead.

Response: Thank you for the suggestion. We have changed our title to 'The effects of health status on work exit and absenteeism among the older working population in China'.

Page 5, line 40: can you define a 'cadre' please?
Response: We have replaced 'cadre' with 'civil servants and employees for state enterprises' on page 5.

Page 7, line 12: What were your hypotheses?

Response: We have added our hypotheses in the Introduction section. Please see page 7.

Page 7, line 28: Some detail of the sampling frame, response rate, geographical area from which the study population was drawn would be helpful here without having to refer to Zhao et al.

Response: As per the suggestion, we have added more detail information on the sampling frame, response rate and geographical area from which the study population was drawn. Please see page 8.

Page 8, line 54: Can you give more details in the main text of how these other health conditions were measured. Did you use standard scales?

Response: Due to the Journal word limitation, we have provided detail information on other health conditions and clarified the definitions of disability condition, the number of chronic diseases and functional limitation in the Appendix. Specifically, in terms of the number of chronic diseases variable, CHARLS included 14 chronic conditions: 1.hypertension; 2.dyslipidemia; 3.diabetes; 4.cancer; 5.chronic lung disease; 6.liver disease; 7.heart problems; 8.stroke; 9.kidney disease; 10.stomach or other digestive disease; 11.emotional, nervous, or psychiatric problems; 12.memory-related disease; 13.arthritis or rheumatism; and 14.asthma. Regarding the functional limitation variable, following Hu's work (Hu et al. 2015), we used 18 items in CHARLS to describe the function limitation of participates. Specifically, functional limitations are assessed in three domains: 7 items measuring physical functions (1.running/jogging about 1 km; 2.getting up from a chair; 3.climbing several flights of stairs; 4.stooping, kneeling or crouching; 5.reaching or extending arms; 6.lifting or carrying over 5 kg; and 7.picking up a small coin), 6 items measuring basic activities of daily living (BADLs) (1.dressing; 2.bathing; 3.eating; 4.getting in/out of bed; 5.using the toilet; and 6.controlling urination and

defecation), and 5 items measuring instrumental ADL (IADLs) (1.doing household chores; 2.preparing hot meals; 3.shopping for groceries; 4.managing money; and 5.taking medications). Each item is measured using a 4-likert scale, “1= No, I don’t have any difficulty”, “2=I have difficulty but can still do it”, “3= Yes, I have difficulty and need help” and “4= I can not do it”. The functional limitations are scored as a total number of items with answers at scale 3 or 4 for functional limitations and at scale > 1 for BADL and ADL. The variable definition table have been provided in the Appendix 1.

The discussion could include more comparison with the results of international studies.
Response: As per the suggestion, we have added more comparison with the results of international studies. Please see page 16.

Reviewer: 2
Reviewer Name: Jan Fekke Ybema
Institution and Country: Utrecht University, The Netherlands
Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below
This is an interesting manuscript about the relationship between self-report health and exit from work and sickness absence among older workers in China. The manuscript reports on a longitudinal study with a baseline measurement in 2011, and a follow-up measurement 2 years later in 2013. It focuses on changes in health status (poor vs. good/fair) and how this is related to exit from work between both measurements and sickness absence as reported at follow-up. It distinguished four groups of workers: male and female farmers and non-farmers. This manuscript seems to be a revised manuscript, however, it is new to me, so I have little choice but to look at it as a new manuscript.

Although the manuscript has a number of interesting features, I also see several serious flaws and limitations that I think should addressed before the manuscript would be ready for publication. The most interesting feature is the attempt to find predictors of exit from work among an understudied group of workers in China. This is highly relevant given the aging population.

I would like to focus on the following problematic aspects of the manuscript that I think need to be changed or addressed: (1) the analytical strategy; (2) the reasons for work exit; (3) the validation of self-reported health; (4) the role of demographic and other control-variables; (5) the term “productivity loss”; (6) the role of sickness absence; (7) details of data collection.

(1) Analytical strategy
The most important point is the analytical strategy. Now, in the central analysis on which conclusions are based, the change in self-reported health (SRH) in 2011 and in 2013 is used to predict exit from work between 2011 and 2013. This means that SRH in 2013 is measured AFTER exit from work. In longitudinal research, independent variables (SRH) should be measured BEFORE the dependent variable (exit). This means that only the results for model I are based on prospective analyses. I think the results for model II (“concurrent health” but actually health following exit), model III, and model IV should be deleted. This simplifies the manuscript strongly, and the resulting analysis is truly prospective.
Response: Thank you for your suggestion. We agree that SRH in 2013 is a measure after exit from work for farmers whose work exit was determined by their working status in the past year. However, the concurrent health issue (health following exit) is not a concern for non-farmers because their work exit was determined by their working status in the last week as well as current status (i.e., were on leave but expected to go back or still received salary). To address the issue for farmers while balancing the analysis for both farmers and non-farmers, we decided to delete model II and model III

as suggested, but still keep model IV (change of health status). The reason we kept model IV is that it helps to understand to what extent the impact of persistently poor health might be different from that of recent health deterioration. We have now focused on either previous health or the trend of health change over time rather than the concurrent health.

It might be interesting to examine SRH at 2013 as a DEPENDENT variable that is regressed on SRH 2011 and on exit between 2011-2013. This would indicate to what extent exit from work lowers or increases SRH. However, this would be the answer to a different research question than is currently examined.
Response: Thank you for your valuable suggestion. We agree that although it is an interesting topic to exam how the change of work status affects people's health, it answers a different research question. We would like to pursue this topic in the future study.

Secondly, I think the separate analyses for males and females, and for the four groups (female farmers, female non-farmers, male farmers, male non-farmers) should ONLY be presented when gender, occupation and their interaction significantly INTERACT with SRH 2011 in the regression of exit from work. If this is not the case, gender and/or occupation (and their interaction) could just be entered as main effects in the regression of exit from work. This would probably be better, given the relatively low numbers of exit from work in some of the groups. My guess would be that there will only be main effects for gender (females have a higher risk of exit) and occupation (non-farmers have a higher risk of exit), in addition to a main effect of SRH 2011 (more exit as health is poorer).
Response: On page 5 of the Introduction section, we have described characteristics of China's labor market, mainly focusing on describing the difference in employment rate between rural and urban areas and provided the possible explanation. Generally speaking, urban population (mainly conducting non-agricultural work) tends to stop working much earlier than the rural counterparts (mainly conducting agricultural work). This divergence of employment for the urban and rural areas is mainly due to the fact that the retirement policy, the pension program and unemployment insurance program are limited only to the urban formal sectors in China but not to the rural population. Therefore, health status is very likely to have different impacts on work exit and sickness leaves for urban and rural people respectively. In addition, previous studies have shown that health status had different impacts on male and female (Bound et al., 1999; Au et al. 2005; Cai & Kalb 2006), therefore, in the present paper, we distinguished four groups according to gender/working types. The sex/gender-disaggregated analysis has been recommended to integrating and evaluating sex/gender in health research (Nowatzki and Grant, 2011; Day et al., 2016; Canada's federal funding agency for health research, the Canadian Institutes of Health Research). Since we expect the difference in work exit and absenteeism among the four groups and so do the impacts of all the independent variables including health status, age, education, etc. among these four groups, we chose to conduct analyses by breaking down the data by gender and working types. We have added more justifications on page 12. To confirm the reviewer's hypothesis, we tested the interaction terms by adjusting for other covariates and they are marginally significant ($0.05<p<0.1$).

Parameter	Level	Estimate	StdErr	tValue	Probt
notwork.hstatus11	1	0.341	0.216	1.575	0.115
notwork.hstatus11	2	0.127	0.155	0.823	0.410
notwork.hstatus11	3	0.000			
notwork.male	1-male	-0.641	0.153	-4.176	0.000
notwork.male	2-female	0.000			

notwork.farm	0-nonfarm	0.081	0.157	0.517	0.605
notwork.farm	1-farm	0.000			
notwork.hs_male	0:poor,male	0.391	0.206	1.899	0.058
notwork.hs_male	1:fair,male	0.117	0.182	0.645	0.519
notwork.hs_male	2:Others	0.000			
notwork.hs_farm	0:poor, nonfarm	0.264	0.263	1.004	0.315
notwork.hs_farm	1:fair, nonfarm	0.320	0.185	1.728	0.084
notwork.hs_farm	2:Others	0.000			
notwork.hs_male_farm	0:poor,male, nonfarm	0.166	0.287	0.577	0.564
notwork.hs_male_farm	1:fair,male, nonfarm	0.265	0.148	1.793	0.073
notwork.hs_male_farm	2;good,male, nonfarm	0.221	0.205	1.076	0.282
notwork.hs_male_farm	3:Others	0.000			

Thirdly, I do not think it is necessary to dichotomize or trichotomize SRH. The full 5-point scale could be used as a “continuous” (or preferably ordinal) scale.
Response: Assuming a 5-point scale as a continuous scale implies that each one unit decrease in SRH was associated with a constant/same impact on work exit. We have tested this assumption and found that this is not the case. The SRH (good or better than good health) had a very similar impact (coefficients) on work exit and the SRH (poor or worth than poor) had a similar impact, but the fair SRH had significantly different impact on work exit from good SRH or poor SRH. In this case, it is reasonable to trichotomize SRH.

Parameter	Level	Estimate	StdErr	tValue	Probt
SRH11	1-very good	-0.745	0.402	-1.852	0.064
SRH11	2 - good	-0.722	0.368	-1.963	0.050
SRH11	3 -fair	-0.326	0.353	-0.923	0.356
SRH11	4 -poor	-0.059	0.367	-0.161	0.872
SRH11	5-very poor	0.000			

Fourthly, I do like trajectory analyses in which changes in SRH are related to the timing of exit from work. However, to do such trajectory analyses in a useful way, you would need more than two measurements. And preferably a much more fine-grained measurement, in which for example each month SRH is measured and changes in SRH before and after the moment of exit from work are analysed using a time series approach.
Response: We agree with the reviewer that a more fine-grained measurement of SRH is preferable in trajectory analysis. Due to lack of data, we are unable to do such analysis at the present time. However, as more data are available, we might be able to do the trajectory analysis in the future.

Finally, it is not fully clear to me what the benefits are of the relatively complex PROBIT analysis over a normal logistic regression, and how the results would be different. I understand that the focal regression (of work exit) is simultaneously done with the validation of SRH. However, I do not understand how this would change the results of either of these analyses. I also think the explanation in the Appendix is now overly technical, and the more interesting results from these analyses are

missing (see below, (3) and (4)).
Response: The probit regression and logistic regression can be used to address the same research question and the conclusions would be the same. In our situation (to construct an index of health), the procedure we used can only handle probit models.

(2) Reasons for work exit
As is mentioned in the Discussion, a minority of the participants who exit from work were not working due to disability or retirement. What other reasons did participant have for not working? I suppose that many of them were unemployed or that there was no work available. We know that unemployment leads to health problems (e.g., the meta-analysis by McKee-Ryan et al., JAP, 2005), which makes it even more important to distinguish health as a predictor and health as a result of exit (see (1)). Is it possible to distinguish voluntary and non-voluntary exit from work? It may be interesting to distinguish several exit routes, as was done in Leijten, et al (2015, your reference 18), using competing risk proportional hazard models.
Response: As shown in the Discussion section (page 18), we did not have enough samples to conduct analyses by distinguishing work exit by different reasons (health-related, early retirement, and others). For example, there were only 6 female non-farmers and 18 male non-farmers exited from work due to health problem and 20 and 9 due to early retirement, respectively. We have described this as one of our limitations in the Discussion section (page 18).

(3) Validation of SRH
In the Appendix an econometric model is specified. If I understand correctly, equations (3) and (4) regard some kind of validation of SRH based on more detailed health measures. However, no results of this validation is given. What percentage of variance of SRH is explained by these detailed measures? Would it be possible to create a more valid and detailed single health measure based on these scales? The items strongly resemble the physical items of SF-12 or SF-36, which are used to create validated scores for mental health and a score for physical health on a scale from 0 to 100. Such a score would seem more reliable and valid than the trichotomized single item on SRH that is used in the analyses.
Response: Following the method of Bound et al. (1999), the equations (3) and (4) in the Appendix 3 were to construct an index of health to address the “justification bias” and measurement error we described in our manuscript (page 11). We have provided a number of goodness-of-fit measures including different Pseudo R-squared for the ordered Probit models for SRH in the Appendix 5. Based on Louviere et al. (2000), “Stated Choice Methods: Analysis and Applications” (page 54), Cambridge University Press, one should not expect to obtain pseudo R^2 values as high as the R^2 commonly obtained in ordinary least squared (OLS) regression applications. For instance, values of McFadden’s LRI between 0.2-0.4 indicate extremely good model fits, which is approximately equivalence to 0.7-0.9 for R^2 from OLS based on a simulation study. Therefore, from our pseudo R^2 values (McFadden’s LRI ranged from 0.09 to 0.18), we could interpret that the detailed measures moderately to strongly explained SRHs (page 19).

(4) Demographic and other control variables
Apparently, in the regression of exit from work several control variables are included (given equation 1 in the Appendix, page 33). I would like to see the results of these control variables. How were age, education, marriage status, and expenditures related to exit from work? Moreover, did any of these variables moderate the relationship between SRH 2011 and work exit?
Response: As per the suggestion, we have reported all the estimated coefficients from our models in the Appendix 6. Our results show that age, education, marriage status, and expenditures are

associated with work exit in at least one of the subgroups. Since they are also highly associated with health status, we included them in our models as potential confounding variables.

(5) The term “productivity loss”
I think the term productivity loss is confusing and should be removed from the title, and perhaps throughout the manuscript. I would advise the authors to use the term “exit from work” instead (and perhaps “sickness absence”, but see next point). Productivity loss is generally reserved to loss of productivity while working (i.e., due to presenteeism) as a complement of sickness absence.
Response: Thank you for the suggestion, we have changed the title to ‘The effects of health status on work exit and absenteeism among the older working population in China’.

(6) The role of sickness absence
I think it is not that interesting to examine the relationships between SRH and sickness absence. Sickness absence is evidently related to lower health (otherwise it would not be sickness absence). This relationship would only be interesting if it is moderated in a meaningful way by other variables. This may be the case for occupation (farmer vs. non-farmer), but this should be tested first. More interesting would be moderation by working conditions (e.g. job demands or job resources that would amplify or buffer the relationship between health and sickness absence). However, as it is presented now, I do not think the inclusion of sickness absence is worthwhile. Moreover, I think the tentative conclusion that for farmers sickness absence is a way to deal with the lack of opportunities for retirement is highly speculative and not based on the results of the analyses.
Response: We have added more justifications to show why it is important to measure the effects of health status on both work exit and absenteeism (page 6). Our study was to emphasize that the poor health leads to not only work exit but also more absent workdays. It is important to understand the consequence of poor health and encourage older working to not only remain in the labour market but also remain productive. When implementing or evaluating the programs to improve health status, both work consequences are important to consider. In the response to above comment, we have explained why we conducted disaggregated analyses by gender and farmer/non-farmer. As suggested, we have tested the interaction and found the impact of poor health status on absenteeism significantly differ by farmers and non-farmers. It will be interesting to explore the effect modification of job demands or job resources on the relationship between health and sickness absence in the future. The reviewer is right that for farmers sickness absence is a way to deal with the lack of opportunities for retirement is not based on the results of the analyses. Therefore, we have revised this statement in the Discussion section. Please see page 17.

(7) Data collection
Finally, I would like some more information about how the data were collected. Were all participants interviewed face-to-face? If not, how did illiterate participants respond?
Response: we have provided more information about how the data were collected. Specifically, the national baseline survey for CHARLS was conducted between June 2011 and March 2012 and the respondents are followed every 2 years, using a face-to-face computer-assisted personal interview. Please see page 8.

Despite these problems that I see in the current manuscript, I would strongly suggest the author to substantially rework the analyses and revise the manuscript. I do believe this could result in a valuable contribution to the literature on exit from work.

Reviewer: 3
Reviewer Name: Emmanuel Aboagye
Institution and Country: Karolinska Institutet, Sweden
Please state any competing interests or state ‘None declared’: None declared.

Please leave your comments for the authors below
No special comments. See attachment
Reviewer’s Evaluation: The effects of health status on work productivity loss among the older working population in China

Comments to authors

The study examines the effects of health status on work productivity loss. It presents estimates of productivity loss (work exit and absenteeism) due to health problems measured as self-rated health (SRH) among older working people in China. Authors conclude that poor lagged health status or poor concurrent health status is expected lead to work exit and more absent workdays. Also persistent poor health status over time is detrimental to the labour participation of the older working population except for female non-farmers. The effects of health status differ by both gender and type of work. Tailored policies to these different working groups are needed.

Title

Change of title suggestion: The effects of health status on work exit and absenteeism among the older working population in China

Work exit or number of absent workdays due to health problems is not necessarily synonymous with productivity loss. Since the study outcomes are work exit and number of absent workdays, it may be advisable to have it in the title.

Response: Thank you for the suggestion. We have changed our title to “The effects of health status on work exit and absenteeism among the older working population in China”.

NB: that the assumption of within-subject changes can also be explained that a high performing worker who becomes less productive because of health problems might still be one of the best among peers in terms of global performance. What about replacement and what about compensatory strategies by colleagues or when the sick absent worker comes back to work?

Response: Compensation or replacement can help reduce the work productivity losses attributable to sick absence. We agree that work exit and number of absent workdays not necessarily end up with productivity loss. We have changed the title and wordings in the manuscript accordingly.

Abstract

1. Work productivity loss was measured in two ways: work exit and the number of absent workdays. This is just a passing comment. I will return to this in the methods and discussion section.

Response: we have replaced ‘work productivity loss’ with ‘work exit and absenteeism’ in Abstract.

2. Health problems/ status: It will be good to mention what health problems or health status is and how it was measured in the methods of the abstract.

Response: As per the suggestion, we have added this information in the Abstract.

Introduction

1. There is a clearly stated aim for the study. It may help that the aim of the study stands alone. Do authors refer to labour force participation and work exit as the same thing? I guess so. You may want to be consistent with the use of the terminologies throughout the manuscript. Please check also the use of labour market activities and continuous working status.

Response: Thank you for the suggestion. We have replaced 'labour force participation' with 'work exit' throughout the manuscript.

2. Lines 12-20 on page 6 can be moved to part of the introduction which tells readers about other specific aims. Specifically, a further aim of the study was to examine the impact of previous health status and concurrent health status as well as the change of health status over time on work exit and the number of absent workdays due to the health problems among the older people who were previously working.

Response: Thank you for the suggestion. Lines 12-20 on page 6 have been moved to the Introduction section.

3. I think it will help to provide a reference for this sentence on line 33-38 on page 5: Most studies to date have focused on either comparing the incremental effects of different chronic diseases on absent workdays or estimating the incremental productivity loss due to different chronic diseases (REF).

Response: As per the suggestion, we have provided references (page 7).

4. Data source belong in the methods: It was mentioned that the study uses longitudinal, individual level data from the China Health and Retirement Longitudinal Study (CHARLS). I think all this belong in the method section.

Response: We have deleted the data source in the Introduction section and added more detail information under the Data and Study Population subsection of Methods. Please see page 8.

Methods

1. Based on the previous comment, I know you have defined productivity loss as work exit and the number of absent workdays but thinking of farmers makes me wonder if it is indeed a loss if a farmer for instance is absent some days. Can the authors clarify why were particularly interested in this classification of agricultural work and also comparing them with non-agricultural work?

Response: As we have mentioned in the Introduction section, the employment rate is very different between rural and urban areas. Specifically, on the one hand, the employment rate among older population in the urban area (mainly conducting non-agricultural work) is very low, it only reaches 40% among people between 50 and 59, and this rate decreases further to about 20% for those aged 60-64. On the other hand, the employment rate among older population in the rural area (mainly conducting agricultural work) is very high, most people still work between 65 and 69, and by the age of 80, the employment rate is still above 20%. This divergence of employment for the urban and rural areas is mainly due to the fact that the retirement policy, the pension program and unemployment insurance program are limited only to the urban formal sectors in China but not to the rural population. Therefore, health status is very likely to have different impacts on work exit and sickness leaves for urban and rural people respectively. It is more likely that health status has more significant impact on sickness leaves for the rural population.

For absent workdays, how were those working in agriculture respond since they may not be expected to be on the farm every day?

Response: In the CHARLS questionnaire, the participants who were conducting agricultural work were first asked for their work time in the past year through a series of questions (how many months did they work, how many days did they work during a normal work month, and how many hours did they usually work during a normal work day in the past year). After these questions, they were asked about the number of absent workdays due to health problems, that is, “How many days of work did you miss last year due to health problems?” We have added the specific question on page 10.

2. Lines 24-31 on page 7 belong to the section where you describe the procedure for analysis. To investigate the potential worker heterogeneity in the effects of health status on work productivity loss, we divided our sample into four separate groups according to gender and working types in 2011: female farmers (i.e., any agricultural work), female non-farmers (nonagricultural work only), male farmers, and male non-farmers.

Response: As per the suggestion, we have moved lines 24-31 on page 7 to the Econometric models subsection. Please see page 12.

3. On line 43-45 on page 7 it is said that Labour force participation status was determined by a series of questions in CHARLS. Can authors give an example or two of the nature of the questions used? I am also curious about what questions they have used to measure the different types of labor participation including household agricultural work, temporary work, own business and unpaid work etc.

Response: The examples of questions in CHARLS to determine work status are as follows and we have included these in the Appendix 2:

FA001: Did you engage in agricultural work (including farming, forestry, fishing, and husbandry for your own family or others) for more than 10 days in the past year?

FA002: Did you work for at least one hour last week? We consider any of the following activities to be work: earn a wage, run your own business and unpaid family business work, et al. Work does not include doing your own housework or doing activities without pay, such as voluntary work.

If respondent said ‘yes’ on either question FA001 or FA002, then she or he was considered as ‘working’. If the answers on both questions were ‘no’, then the respondent needed to answer the following two questions:

FA005: Do you expect to go back to this job at a definite time in the future or within 6 months?

FA006: Do you still receive any salary from this job?

If respondent said ‘yes’ on either question FA005 or FA006, then she or he was considered as ‘working’.

If respondent said ‘no’ on all four questions, she or he was considered as ‘not working’.

4. Same concern here as in comment 3. Please check line 54 on page 8. These measures included disability condition, number of chronic diseases, and functional limitations. Please provide examples of the questions asked in the appendix or rather still use in-text references to the detailed health questions the first time they are mentioned so that readers can check what kind of question were asked.

Response: we have provided detail information on other health conditions and clarified the definitions of disability condition, the number of chronic diseases and functional limitation in the Appendix. Specifically, in terms of the number of chronic diseases variable, CHARLS included 14 chronic conditions: 1.hypertension; 2.dyslipidemia; 3.diabetes; 4.cancer; 5.chronic lung disease; 6.liver disease; 7.heart problems; 8.stroke; 9.kidney disease; 10.stomach or other digestive disease; 11.emotional, nervous, or psychiatric problems; 12.memory-related disease; 13.arthritis or rheumatism; and 14.asthma. Regarding the functional limitation variable, following Hu’s work (Hu et al. 2015), we used 18 items in CHARLS to describe the function limitation of participates. Specifically, functional limitations are assessed in three domains: 7 items measuring physical functions (1.running/jogging about 1 km; 2.getting up from a chair; 3.climbing several flights of stairs; 4.stooping, kneeling or crouching; 5.reaching or extending arms; 6.lifting or carrying over 5 kg; and 7.picking up a small coin), 6 items measuring basic activities of daily living (BADLs) (1.dressing; 2.bathing; 3.eating; 4.getting in/out of bed; 5.using the toilet; and 6.controlling urination and defecation), and 5 items measuring instrumental ADL (IADLs) (1.doing household chores; 2.preparing hot meals; 3.shopping for groceries; 4.managing money; and 5.taking medications) (Hu et al., 2015). Each item is measured using a 4-likert scale, “1= No, I don’t have any difficulty”, “2=I have difficulty but can still do it”, “3= Yes, I have difficulty and need help” and “4= I can not do it”. The functional limitations are scored as a total number of items with answers at scale 3 or 4 for functional limitations and at scale > 1 for BADL and ADL. The variable definition table have been provided in the Appendix 1.

5. Can authors provide any ethics statement? State, if it is perhaps not applicable in this case.

Response: Ethics approval for this study was not required because it was based exclusively on the publicly available data, CHARLS, and the study subjects were not directly approached. We have added this statement in the Data and Methods section. Please see page 9.

Results and discussion

1. I think the results and discussion sections address the research questions and cover many grounds including the sensitivity analysis performed, and in conclusion where policy measures may be necessary.

Response: We have provided more policy implications in Discussion section. Please see pages 19-20.

2. It seems that female non farmers clearly stand out in terms of change of health status over time on the outcomes - work exit and the average number of absent workdays due to health problems.

What could be the explanation for this? I think the authors can elaborate on this in the discussion beyond the fact that there were only a few female non farmers.

Response: The reviewer is right that female non-farmers stand out in our analysis. We have elaborated more on this finding and added policy implications in the Discussion section (on pages 17-20). This finding suggests that health status is not associated with work exit of female non-farmers. The work exit of female non-farmers is attributable to factors other than health. The improvement in health status only might not keep female non-farmers at work.

Appendix

2. Econometric model specifications

We used the Model III, including both lagged and concurrent health status independently he model, as an example to show our model specifications. Please check line 5 on page 33. There is some text missing where I have marked.
Response: Thank you for pointing out the typo, we have corrected it.

Reviewer: 4
Reviewer Name: Zhiyuan Hou
Institution and Country: Fudan University, China
Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below
This paper addresses an important question. It contributes to literature by estimating the effects of health status on work productivity among the older working population in China, and the results are presented clearly. However, there are several technical details I prefer be revised.
First, the paper uses work exit or the number of absent workdays to measure productivity loss. However, in literature, productivity is often measured as TFP or output related variables such as output per worker.

Response: Thank you for your suggestion. We have removed ‘productivity loss’ and used ‘work exit’ and ‘the number of absent workdays’ instead throughout the manuscript.

Second, analyses in this paper are based upon regression results of 4 separate groups. However, estimates by groups do not provide us evidence about the significance of the difference between two groups. Another way to show heterogeneous effects of health status on productivity for different groups is to use interaction terms in regressions. For example, you can interact gender with health to

show if health has significantly different effects on work exit. Please explain why you choose regression for separate groups instead of interaction effect.

Response: Please see our response to the same comment made by Reviewer 2, comment #2 above.

Third, though instruments are used to estimate the model, regression results suffer from endogeneity issues. The assumption that error terms in the first stage and the second stage are not correlated is very strong. For example, city/regional level characteristics such as economic growth or labor market demand shocks may influence residents' health status and probability of work exit simultaneously, resulting in biased estimates. More city/regional level explanatory variables related to economic/labor market conditions should be included in regressions.

Response: Thank you for your suggestion. To solve endogeneity problem, first, based on reviewer 2's suggestion, we have only kept the original models I and IV, which investigated the impact of health status of 2011 (health status before work exit) and the change of health status between 2011 and 2013 on work exit and the number of absent workdays due to the health problems in 2013, respectively. Second, instead of the method of instrumental variables, we followed Bound et al. (1999) and used the latent variable model to use disability condition, number of chronic diseases, and functional limitations to construct an index of health to address the endogeneity and measurement error issues of the SRH. Please see pages 11-12.

Besides, people may exit the labor market due to different reasons. On the one hand, people may exit the job market voluntarily due to bad health conditions. On the other hand, people may become unemployed (or exit the labor market involuntarily) due to negative demand shocks in the labor market. This paper uses work exit as the dependent variable for both situations, making the interpretations of the results quite confusing.

Response: As shown in the Discussion section (page 18), we did not have enough samples to conduct analyses by distinguishing work exit by different reasons (health-related, early retirement, and others). For example, there were only 6 female non-farmers and 18 male non-farmers exited from work due to health and 20 and 9 due to early retirement, respectively. We have described this as one of our limitations in the Discussion section (page 18).

Fourth, I suggest the authors show how probability of not working, or the expected number of absent workdays are computed in Table 3 to Table 5. It is true that estimated parameters measure the effect of independent variables on latent variables instead of actual dependent variables. Partial effects can always be calculated using maximum likelihood estimations. Please provide marginal effects of health status on work exit and absent workdays.

Response: We have explained how we calculated the expected probability of not working and the expected number of absent workdays for each category of health status or each category of the change of health status (on pages 12-13). The marginal effect is basically the difference in the expected values between any two categories.

Finally, the paper is policy oriented and the question investigated is very important. I hope more policy discussions could be provided.

Response: We have provided more policy discussions. Please see pages 19-20.

Reviewer: 5
Reviewer Name: Cathy Gong
Institution and Country: Australian National University
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below
Review comments on BMJ Open-2018-024115

Title: The effects of health status on work productivity loss among the older working population in China
Suggestion: Major revision

Summary:

This paper used the first two-wave longitudinal data from the China Health and Retirement Longitudinal Study (CHARLS) to analyze the effects of health status and its changes on work productivity loss among the older working population in China. It measured the impact of self-reported health on work exit and the number of absent workdays, for four groups according to gender and work types (farmers who conducted any agricultural work at the first wave versus non-farmers who conducted non-agricultural work only).

Strength

This paper has filled in the research gap to address a very important policy issue on how to encourage older people to work as long as they can in order to face the emerging workforce shortages due to the fast population ageing by improving older people's health.

The use of measures of work exit and absent days from work, previous health status, current health status and health changes across two waves are very `interesting and meaningful.

It is expected that persistent poor health over time has larger impact on work exit and absence days of work.

Limitations:

(1) The statistical methods used are not enough to support the research question, which need to be improved.

(2) The interpretation of stories by gender, famers/non-farmers is not deep enough.

Major comments:

(1) How is the quality of instrumental variables (IV) (IV=disability condition, number of chronic diseases, and functional limitations) used to address possible endogeneity between self-reported health and work exit or absence days from work? Can you say the three IV variables had direct impact on self-reported health but not on work exit and absent workdays?

Response: To clarify the criteria of IV, IV needs to be 1) associated with the main determinant (SRH) but 2) unrelated to the error term for the main effect, i.e., unrelated to the outcome(s), other than through its association with SRH. That is, the impact of functional limitations e.g., on work exit and sickness leave is assumed to go through its association with SRH, i.e., the SRH is on the impact

pathway of functional limitation on work exit and sickness leave. Most studies (e.g., Van den Berg et al Occup Environ Med 2010, Van Rijn et al J Occup Environ Med 2014, Leijten et al J Epidemiol Comm Health 2015) measured the impact of functional limitations without adjusting for SRH, i.e., functional limitations and SRH were not included in the same regression model. However, based on the study from Bound et al., 1999, it is more appropriate to describe our approach as "latent variable model", which is analogous to using the three detailed health measures to construct an index of health. In addition, we have presented the goodness-of-fit measures to indicate how well the three detailed health measures predict the SRH. Please see page 19 and Goodness-of-fit measures section in the Appendix 5.

(2) In you case, as you have two waves of longitudinal data, you might be able to just use difference-in-difference methods or conditional difference-in-difference methods to control for the endogeneity between work and health, instead of IV method.
Response: Thank you for your suggestion. To solve endogeneity problem, first, based on reviewer 2's suggestion, we have only kept models I and IV, which investigated the impact of health status of 2011 (i.e., the health status before work exit) and the change of health status between 2011 and 2013 on work exit and the number of absent workdays due to the health problems in 2013, respectively. Second, we followed Bound et. al.(1999) and used the latent variable model to use disability condition, number of chronic diseases, and functional limitations to construct an index of health to address the endogeneity and measurement error issues of the SRH. In addition, our study design is similar to difference-in-difference methods. We measured the impact of the difference of health status over time (poor to poor, poor to good, good to poor, and good to good) on the one-way difference of work status over time (work to not working or work exit). The work exit among people with good to good health status change reflected the natural work exit trend. Thus, for example, the difference in work exit between people with good to poor and people with good to good represented the additional impact of poor health.

Other comments in details:

(1) As previous and current health status are highly correlated, you should use either initial health plus health changes, or current health plus health changes in the models of Tables 3 and 4 , but not using previous and current health status in Table 3 or health changes only in Table 4.

Response: As suggested by Reviewer 2, we have kept two main models: one investigated the impact of previous health status; the other investigated the impact of the change of health status. Our change of health status was defined as poor to poor, poor to good, good to poor and good to good, which combines two variables, initial health in 2011 (poor or good) plus health changes (better, same or worse), into one variable.

(2) You need provide data source and notes for all the tables;

Response: Thank you for the suggestion. We have provided more detail information on data source and how it were collected. Please see page 8.

(3) It is important to check whether using weights or not generated robustness results, but there is no need to report all the tables in the appendix. Instead, I would like you to report all the estimated coefficients from your models on self-reported health, work exit and absent workdays in the appendix.

Response: As per the suggestion, we have reported all the estimated coefficients from our model in the Appendix.

(4) There are a lot of important differences between farmers and non-farmers, males and females in Table 1 and Table 2, which are very interesting, but you have not discussed them. For instance, non-farmers have much better health than farmers, while farmers work longer to late life than non-farmers; changes from poor to poor health has large impact for female non-farmers, male farmers, male non-farmers, while changes from good/fair to poor health has large impact on female farmers, is this because female farmers take more role in unpaid domestic work instead of paid agriculture work? Why worsening health has larger impact on work exit for male non-farmers than female non-farmers, is this because relatively later retirement age for males? Should females have same retirement age with males from health view?

Response: Thank you for your suggestion. We have discussed the gender difference and difference between farmers and non-farmers in terms of their health status, work exit and absenteeism in the Discussion section. We have also added the corresponding policy implications such as the extension of female legal retirement as suggested. Please see all the changes on pages 17, 18, 19 and 20.

(5) Will your findings indicate that retirement age could be be postponed to later age in urban area from health view? Will work flexibility (like in rural area) help older people to work longer by taking more unpaid leave while staying in workforce longer?

Response: Our study does not provide direct evidence to suggest whether retirement age could be postponed to later age in urban area from health view. On the contrary, our findings of older farmers taking more sick leaves to remain in the labour force might suggest an unproductive rural labour force. It may indicate that the new pension plan or disability benefit has not provided enough social security for the elderly in rural China or there is a lack of knowledge and awareness of such plan or benefit.

(6) You have checked the robustness of your results by removing the age restriction for farmers, while it will be more interesting to compare the impact of health changes for farmers before and after legal retirement age. This could be used to inform non-farmers behaviours if the retirement age is cancelled.

Response: Thank you for your suggestion. We agree that it is interesting to compare the impact of health changes for farmers before and after legal retirement age or to identify which age shall be set as the retirement age. We think this can be a different research question and would like to pursue this topic in the future study. Our initial exploration shows the health has more impact on work exit among farmers after legal retirement age (i.e., the older population), which is expected. However, we would not say that the research findings can be used to inform non-farmers' behaviours without further research, considering the difference between rural and urban areas, for example, different job characteristics and different social security policies between farmers and non-farmers, etc.

(7) Do you know how many people are still working after retirement age for farmers and non-farmers?

Response: Please see the number reported in the Table below.

year	farmer	retirement_age	number	percent
2011	Non-farmers	Under	1950	78.47
2011	Non-farmers	Over	535	21.53
2011	Farmers	Under	4498	50.43
2011	Farmers	Over	4421	49.57
2013	Non-farmers	Under	2073	74.97
2013	Non-farmers	Over	692	25.03
2013	Farmers	Under	4248	46.21
2013	Farmers	Over	4945	53.79

VERSION 2 – REVIEW

REVIEWER	Jan Fekke Ybema Utrecht University, the Netherlands
REVIEW RETURNED	13-Mar-2019

GENERAL COMMENTS	Unfortunately, the revision of this manuscript failed to convince me. The authors followed some of my suggestions, but not others. I regard their choice not to test whether the results are different in the four groups of participants as problematic as their main conclusions are based on observed (but untested) differences. Examining the regression weights for Model 1 in Table 3 and the proportions of exit from work in Table 4 seem to suggest that for female non-farmers the effects of poor health in 2011 on work exit are similar as for female and male farmers. Only for male non-farmers health in 2011 seems unrelated to work exit. I would still prefer pooling of the four groups and testing differences with interactions to prevent overinterpretation of differences in statistical significance rather than substantial differences in effect size. In their response to my review, the authors present results in which they test the requested interaction effects. It is not fully clear to me how they did this or how to interpret the parameters. Moreover, if I understand them correctly, these results do not seem in line with the results presented in the article, as the main effects of health apparently are not significant, nor are the main effects of farmer/non-farmer. This is perplexing given the relatively large differences in overall percentages in work exit for different levels of health and for farmers versus non-farmers. Moreover, I would like to see a more comprehensive description of the main results in the article. The authors describe the main result in only one sentence, which is rather unclear, i.e., “Model I showed that people with poorer lagged health status except for non-

	farmers were significantly more likely to exit from work." (page 14). They rely on the readers to closely examine the complex tables of results themselves without providing adequate explanation of these results. For example, it is not clear to me what the numbers in Table 2 present for "proportion of work exit". These are not proportions, but seem to be numbers of participants who exit work. Are the numbers in parentheses then weighted percentages (rather than proportions)? As I mentioned in my previous review, I do not think the Model with change in health status is very well interpretable as the health measure follows the work exit. I do not understand the authors' claim that work exit is not preceding the 2013 measure of health in non-farmers. They suggest that work exit among non-farmers means that they "… were on leave but expected to go back or still received salary" (authors' response to my review, point (1)), but in the Method section this group of participants are labeled as "working" rather than "not working" (page 9). Finally, I continue to find the chosen analytical strategy too complex. I am not convinced of the added value of validating the SRH-measure with the more detailed health measures. It remains fully unclear how this validation would influence the main results that are presented in the manuscript. An alternative strategy could be to use structural equation modeling with a multigroup approach. This would also make it possible to test differences in relationships between the groups, and a latent variable for health could be constructed based on several related indicators. Detailed comments: • Page 9 "and the study subjects were not directly approached". This is unclear. I suppose the authors mean that only secondary analyses are carried out and no new data were collected? However, ethical approvals would seem necessary for CHARLS as a whole, and whether this ethical approval was obtained is the relevant question. • Page 11 "There are no plans about dissemination of the results". The present article is a form of dissemination of the results. So please reformulate or explain what is meant with this statement. • Page 12 "All analyses were weighted using the individual longitudinal weights provided by CHARLS". Please explain what these weight do. Do they weight back to the original sample (to control for selective attrition?), to the older Chinese working population? Or something else. • Page 14 "Consistently among all the four groups, people in poor health status in 2011 or 2013 were more likely to stop working in 2013 except for female non-farmers". I do not see from Table 2 how this can be concluded. It seems (if the values in parentheses are percentages) that work exit is least influenced by 2011 health in female farmers rather than female non-farmers. • Page 18 I would present the percentages about not working due to health reasons or retirement in the four groups in a Table rather than in the text. It is now rather hard to follow. I also think that these additional results could better be placed in the Results section rather than the Discussion. • Page 19 I also think the description of the pseudo R2 had better be placed in the Results section together with a paragraph on the validation of the health measure (if this validation is retained in the manuscript). It does not add much to the manuscript as it is now.

	• Page 21 “Poor two-year lagged health leads to work exit of female and male farmers”. I would change “leads” to “predicts”, to prevent unwarranted causal claims. • Page 21 “Persistently poor health or recent health deterioration over time is most detrimental among all older Chinese workers except female non-farmers”. Detrimental for what? As it is now, I do not think this statement is valid or informative. Almost by definition persistent poor health or recent health deterioration is detrimental for everyone.

REVIEWER	Zhiyuan Hou Fudan University, China
REVIEW RETURNED	28-Mar-2019

GENERAL COMMENTS	all my comments have been solved.

REVIEWER	Cathy Gong Australian National Univesity
REVIEW RETURNED	13-Mar-2019

GENERAL COMMENTS	The revised version of this paper has been improved greatly after carefully addressing the reviewers’ comments. I don’t have further comments. From my view, it is acceptable for publication now.

VERSION 2 – AUTHOR RESPONSE

Reviewers' Comments to Author:

Reviewer: 2

Reviewer Name: Jan Fekke Ybema

Institution and Country: Utrecht University, the Netherlands

Please state any competing interests or state ‘None declared’: None declared

Unfortunately, the revision of this manuscript failed to convince me. The authors followed some of my suggestions, but not others. I regard their choice not to test whether the results are different in the four groups of participants as problematic as their main conclusions are based on observed (but untested) differences. Examining the regression weights for Model 1 in Table 3 and the proportions of exit from work in Table 4 seem to suggest that for female non-farmers the effects of poor health in 2011 on work exit are similar as for female and male farmers. Only for male non-farmers health in 2011 seems unrelated to work exit. I would still prefer pooling of the four groups and testing differences with interactions to prevent overinterpretation of differences in statistical significance rather than substantial differences in effect size.

Response: Thank you for your suggestion. Our objective is to analyze the effects of health status on work exit and absenteeism among the older working population in China. Due to the differences between males and females and the differences between farmers and non-farmers in China, as we mentioned in our previous response, we have strong justifications to examine the effects of health status by separating entire population into four groups. These justifications were based on background knowledge rather than statistical testing or data-driven. In addition, not only the impact of health status but also the impact of other independent variables on work exit/absenteeism may differ between the groups. If we pooled the four groups together and only included the interaction terms among the four groups and health status, we would assume that the impacts of other independent variables on work exit/absenteeism were the same among the four groups, which is not reasonable. However, it would not be realistic to include interaction terms among all independent variables with gender and work type. Therefore, we will insist on using separate models for our analyses. Nevertheless, as suggested by the reviewer in his previous comments, we had conducted a preliminary analysis by pooling all four groups and testing difference with interactions. It shows that, conditional on the same impacts of other independent variables on work exit among the four groups, the impacts of health status on work exit marginally differ among the groups, which confirms that four separate group analyses are needed. We have added the test results in the Section 4 of Appendix.

What we have done in the manuscript was to compare the impacts of different health status on work exit/absenteeism within each subgroup instead of the comparison across the subgroups. To avoid the possible confusion, we now present the effects by groups and delete the sentence 'The effects of health status differ by both gender and work type' from the conclusion paragraph in both Abstract and Discussion sections. Please see pages 2-3, and 21.

In their response to my review, the authors present results in which they test the requested interaction effects. It is not fully clear to me how they did this or how to interpret the parameters. Moreover, if I understand them correctly, these results do not seem in line with the results presented in the article, as the main effects of health apparently are not significant, nor are the main effects of farmer/non-farmer. This is perplexing given the relatively large differences in overall percentages in work exit for different levels of health and for farmers versus non-farmers.

Response: SAS QLIM procedure was used to carry out the analysis. We included the interactions between health status and work type, health status and gender, and health status and work type and gender in the model. The significances of the interaction parameters indicate whether the impact of health status on work exit differ among the groups. The results were not expected to be the same as those presented in our paper, because they were from different models. In the model with interaction terms, it was assumed that the impacts of other independent variables, such as age, education, marriage status, and household expenditures, on work exit were the same among the four groups. However, by using separate models, we allowed the difference in the impacts of all independent variables on work exit among the four groups. Considering the differences among the four groups in work exit/absenteeism, health status, and all other characters, we chose using separate models for the four groups of people to investigate the impact of health status on work exit/absenteeism.

Compared to the models without interaction terms, the interpretation of parameters from a model with interaction terms is more complicated. For instance, the p-value (0.058) for the parameter of interaction between gender and health status, (the one with level = '0-poor, male' in the table we provided to the reviewer, now in Appendix Section 4), indicates that the impacts of health status (Poor vs Good) on work exit marginally differ between male and female farmers (the reference group for

work type). Similarly, the p-value (0.084) for the parameter of interaction between work type and health status (the one with level = '1-fair, nonfarm') indicates that the impacts of health status (Fair vs Good) on work exit marginally differ between female (the reference group for gender) farmers and female non-farmers. The p-values for model parameter of health status, level = '1-poor' or '2-fair', only indicate the significance of the impact of health status (Poor vs Good, Fair vs Good) on work exit among female (the reference group for gender) farmers (the reference group for work type). However, it cannot be concluded that the main effects of health status are not significant from these p-values, because the impact of health status on work exit differ among the four groups and the model doesn't provide a single main effects of health status among the whole population. Please also notice that these significances for the impact of health status on work exit in female farmers, for example, may be different from those presented in our manuscript, because they are from the models with different assumptions. In the same way, the significance of work type parameter (level = 0-nonfarm) only indicates that the impact of non-farmer on work exit was not significant for the females (the reference group for gender) with Good health status (the reference group for health status), but not the main effect of work type on work exit for the whole population. To look at the main effect of an independent variable, such as health status, gender, and work type, among the whole population, one should try a model without the interaction terms of the variable with any other independent variables.

Moreover, I would like to see a more comprehensive description of the main results in the article. The authors describe the main result in only one sentence, which is rather unclear, i.e., "Model I showed that people with poorer lagged health status except for non-farmers were significantly more likely to exit from work." (page 14). They rely on the readers to closely examine the complex tables of results themselves without providing adequate explanation of these results. For example, it is not clear to me what the numbers in Table 2 present for "proportion of work exit". These are not proportions, but seem to be numbers of participants who exit work. Are the numbers in parentheses then weighted percentages (rather than proportions)?

Response: As per the suggestion, we have added a more comprehensive description of the main results in the manuscript. Please see page 15. We have changed the title to 'work exit' in Table 2. The numbers in parentheses are weighted percentages. We have made corresponding changes in Table 2 and added notes at the bottom of Table 2.

As I mentioned in my previous review, I do not think the Model with change in health status is very well interpretable as the health measure follows the work exit. I do not understand the authors' claim that work exit is not preceding the 2013 measure of health in non-farmers. They suggest that work exit among non-farmers means that they "… were on leave but expected to go back or still received salary" (authors' response to my review, point (1)), but in the Method section this group of participants are labeled as "working" rather than "not working" (page 9).

Response: Sorry for the confusion. We had agreed with the reviewer and removed the model with health status in 2013. In our previous response, we mentioned 'for non-famers their work exit was determined by their working status in the last week as well as current status (i.e., were on leave but expected to go back or still received salary)', which is different from how the work exit of famers was defined. Those who were on leave but expected to go back or still received salary were defined as 'working'. Indeed, we were trying to explain that, compared with famers whose work exit were defined as not working for more than 10 days in the past year, work exit for non-farmers were defined as working less than one hour in the last week and not expecting to go back to work. Therefore, the non-farmers' health status and work status could be treated as being measured at the same time (within 7 days). We apologize for not having made this point clear in our previous response. The reason we keep the model with the change of health status is that it helps to understand to what extent the

impact of persistently poor health might be different from that of recent health deterioration but nothing else.

Finally, I continue to find the chosen analytical strategy too complex. I am not convinced of the added value of validating the SRH-measure with the more detailed health measures. It remains fully unclear how this validation would influence the main results that are presented in the manuscript. An alternative strategy could be to use structural equation modeling with a multigroup approach. This would also make it possible to test differences in relationships between the groups, and a latent variable for health could be constructed based on several related indicators.

Response: Thank you for your suggestion. Basically, our model belongs to simultaneous equations models in econometrics. Constructing a health index by more detailed health measures is a well-established method in labour economics literature (Bound et. al. 1999; Disney et. al., 2006). We also believe that the previous reviewers had agreed with our analysis method. Structural equation modeling with a multigroup approach may be an alternative. We may try it in our further studies in this field.

Detailed comments:

- Page 9 "and the study subjects were not directly approached". This is unclear. I suppose the authors mean that only secondary analyses are carried out and no new data were collected? However, ethical approvals would seem necessary for CHARLS as a whole, and whether this ethical approval was obtained is the relevant question.

Response: Sorry for the confusion. The reviewer is correct that only secondary analyses were carried out and no new data were collected for our study. CHARLS was approved by the Ethical Review Committee (IRB) at Peking University, Beijing, China. We have revised this paragraph in the manuscript. Please see page 9.

- Page 11 "There are no plans about dissemination of the results". The present article is a form of dissemination of the results. So please reformulate or explain what is meant with this statement.

Response: Sorry for the confusion. The reviewer is correct that the manuscript is a form of dissemination of the results. What we claimed in the manuscript was that this article will not be disseminated to the participants of CHARLS particularly. To remove this confusion, we have removed this sentence from the manuscript. Please see page 11.

- Page 12 "All analyses were weighted using the individual longitudinal weights provided by CHARLS". Please explain what these weight do. Do they weight back to the original sample (to control for selective attrition?), to the older Chinese working population? Or something else.

Response: Sorry for the confusion. These are sampling weights constructed by CHARLS research team for the longitudinal respondents (interviewed in both 2011 and 2013) to represent the older population in China. We now have added the reference to the manuscript (ref: http://charls.pku.edu.cn/uploads/document/2013-charls-wave2/application/CHARLS_Wave2_Release_Note.pdf) for the interested readers.

- Page 14 “Consistently among all the four groups, people in poor health status in 2011 or 2013 were more likely to stop working in 2013 except for female non-farmers”. I do not see from Table 2 how this can be concluded. It seems (if the values in parentheses are percentages) that work exit is least influenced by 2011 health in female farmers rather than female non-farmers.

Response: Sorry for the confusion. What we have done in the manuscript is to compare the impacts of different health status on work exit within each subgroup instead of the comparison across the subgroups. For example, among male farmers, the percentages of work exit in 2013 were 3.76% and 4.30% for those in good and fair health status in 2011, respectively; and this number became 8.40% for those in poor health status. Those numbers indicated a possible association between the health status in 2011 and the probability of work exit in 2013 among male farmers. To eliminate the confusion, we have reframed our statements in the manuscript. Please see page 14.

- Page 18 I would present the percentages about not working due to health reasons or retirement in the four groups in a Table rather than in the text. It is now rather hard to follow. I also think that these additional results could better be placed in the Results section rather than the Discussion.

Response: Thank you for the suggestion. We now reported the percentages about not working due to health reasons or retirement in a Table and put it in Appendix. We decided to keep this information in the Discussion section since although these were parts of the results, they were not our main focuses. The main reason we reported them was to discuss the limitation. Therefore, we felt it would be more appropriate to put them in the Discussion section.

- Page 19 I also think the description of the pseudo R2 had better be placed in the Results section together with a paragraph on the validation of the health measure (if this validation is retained in the manuscript). It does not add much to the manuscript as it is now.

Response: As per the suggestion, we have moved the description of the pseudo R2 to the Results section and moved the corresponding table from Appendix to the main manuscript. Please see pages 14-15.

- Page 21 “Poor two-year lagged health leads to work exit of female and male farmers”. I would change “leads” to “predicts”, to prevent unwarranted causal claims.

Response: Thank you for the suggestion. We have changed “leads” to “predicts” in the manuscript.

- Page 21 “Persistently poor health or recent health deterioration over time is most detrimental among all older Chinese workers except female non-farmers”. Detrimental for what? As it is now, I do not think this statement is valid or informative. Almost by definition persistent poor health or recent health deterioration is detrimental for everyone.

Response: Thank you for the suggestion. We have changed the sentence to: Persistently poor health or recent health deterioration over time has detrimental impact on labour market in terms of work exit and absenteeism among all older Chinese workers except for female non-farmers. Please see page 21.